DOCUMENT RESUME

ED 302 582 TM 012 649

AUTHOR Hambleton, Ronald K.; Rogers, H. Jane
TITLE Design of an Item Bias Review Form: Issues and Questions.
SPONS AGENCY New York State Education Dept., Albany.
PUB DATE 88
NOTE 19p.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Content Analysis; Culture Fair Tests; Ethnic Stereotypes; *Evaluation Methods; Item Analysis; Language Patterns; Minority Groups; Religious Factors; Sex Stereotypes; *Stereotypes; *Test Bias; Test Construction; Testing Problems; Test Items

ABSTRACT

Issues in preparing a review form to detect item bias in tests are discussed and the first draft of an item bias review form is presented. While stereotyping is the consistent representation of a given group in a particular light, bias is the presence of some characteristic of an item that results in differential performance of two individuals of equal ability but from different subgroups. Both are undesirable properties of a test. Stereotyping and inadequate representation are discussed as they apply to tests. Bias may involve: (1) sex, cultural, ethnic, class, and religious factors; (2) content; (3) language; (4) item structure and format; and (5) test time limits. The sample bias review form was designed to assist in the identification of items that may reflect bias against designated subgroups of interest. The respondent answers 19 questions about bias for each test item and answers 4 questions of judgment on the test as a whole. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Ronald Hambleton
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED302582

Design of an Item Bias Review Form: Issues and Questions[1]

Ronald K. Hambleton and H. Jane Rogers
University of Massachusetts at Amherst

The purposes of this report are (1) to provide a brief discussion of the issues that arise in preparing a judgmental review form for detecting item bias; and (2) to offer a first draft of an item bias review form for discussion and subsequent revision. A list of useful references is included at the end of the report.

Definitions

In any investigation of bias, the first step is to identify the subgroups of interest. Bias reviews and studies generally focus on differential performance for sex, ethnic, cultural, and religious groups. For the purposes of New York state, some or all of these groups may be of interest. In the discussion which follows, we refer to "designated subgroups of interest," or "DSI," in order to avoid repeating a list of possible subgroups. Once the state has identified which groups are of interest, more specific terms can be substituted.

An important distinction is made in the following discussion between stereotyping and bias. Stereotyping is the consistent representation of a given group in a particular light, which may be offensive to members of that group. Stereotyping does not, except in extreme cases, lead to differential performance between the designated subgroups of interest (DSI). Bias, on the other hand, is the presence of some characteristic of an item which results in differential

[1] This work was supported by the New York State Department of Education.

TM012649

performance for two individuals of the same ability but from different subgroups. Both stereotyping and bias are undesirable properties of a test item, and, hence, both issues are addressed below and on the review form.

Issues in Preparing an Item Bias Review Form

Fairness vs. Bias

In preparing an item bias review form, an important consideration is that of the direction of the questions on the form. Each question can be asked from two perspectives: Is the item fair?, or, Is the item biased? While the difference may seem trivial, some researchers contend that judges cannot detect bias in an item, but can assess the fairness of an item. Perhaps the best approach to take is to include questions of both types on the review form. A list of questions which address fairness is given below. Questions concerning bias are presented in the following sections.

- Does the item appear to be fair with respect to representation of situations for examinees, and free of annoying stereotyping?
- Does the item give a positive representation of DSI?
- Is there a lack of representation of DSI in non-stereotypical settings?
- Is the test item material balanced in terms of being equally familiar to every DSI?
- Is there an over- or under-representation of a sex group in either a primary or a secondary role?

- Are members of DSI highly visible and positively portrayed in a wide range of traditional and non-traditional roles?
- Are positive stereotypes (example: A woman as a loving mother) balanced by a sufficient number of nontraditional portrayals?
- Are DSI represented at least in proportion to their incidence in the general population?
- Does the item include topics of interest and relevance to DSI?
- Are DSI referred to in the same way with respect to the use of first names and titles?
- Is there an equal balance (across items in the test) of:

  proper names?

  ethnic groups?

  activities for all groups (active, passive neutral)?

  roles for both sexes (traditional, nontraditional, neutral)?

  adult role models (worker, parent)?

  character development (major, minor, neutral)?

  settings (city, suburban, urban, rural, neutral)?
- Does an item have contextual justification? (example: predominance of sickle cell anemia among Black people)
- Is there greater opportunity on the part of members of one group to be acquainted with the vocabulary?
- Is there greater opportunity on the part of members of one group to experience the situation or become acquainted with the process presented by the items?
- Will the item "turn-off" examinees so that they are unable to do as well as their abilities would indicate?

- Will all examinees be "free" psychologically and emotionally to respond to an item?
- Will all examinees have equal opportunity to respond?
- Are the members of a DSI portrayed as uniformly having certain aptitudes, interests, occupations, or personality characteristics?

Stereotyping and Inadequate Representation of Minorities

Stereotyping and inadequate or unfavorable representation of DSI are undesirable properties of the test to which judges should be sensitized. The test should be free of material which may be offensive, demeaning, or emotionally charged for some groups. While the presence of such material may not make the item more difficult for the candidate, it may cause him or her to become "turned off", and result in lowered performance. An example of emotionally charged material would be an item dealing with the high suicide rate among Native Americans. An example of offensive material could be that of an item which implied the inferiority of a certain group would be offensive to that group. Terms which are generally unacceptable in test items include: lower class, housewife, Chinaman, colored people, Redman.

Questions which address the issue of stereotyping and inadequate or unfavorable representation of minorities are listed below.

- Does the test item contain material which is controversial or inflammatory for DSI?
- Does the item contain material which is demeaning or offensive to members of DSI?

- Does the test item portray members of DSI in situations that do not involve authority or leadership?
- Does the test item depict members of either sex as experiencing stereotyped emotions (example: girls crying, boys being brave)?
- Does the test item depict members of DSI as having stereotyped characteristics (example: blacks as poor people)?
- Does the test item depict members of DSI in stereotyped occupations (example: Chinese launderer)?
- Does the test item depict members of DSI in stereotyped situations (example: boys as creative and successful, girls needing help with problems)?
- Does the test item contain "art bias" (example: girls always in dresses and ribbons)?
- Does the item contain language that could be offensive to a segment of the examinee population?
- Does the item contain biased language? (For example, disproportional uses of male terms or names and patronizing expressions like "the little woman" or "the fair sex" must be avoided.)
- Do the job designations end in "man" (example: use police officer instead of policeman; use firefighter instead of fireman).
- Have offensive terms such as man, men, and mankind, been used as collective terms for the human race? (Instead, use such terms as humanity, people, men and women.)

Sex, Cultural, Ethnic, Religious, and Class Bias

An item may be biased if it contains content or language that is differentially familiar to subgroups of examinees, or if the item structure or format is differentially difficult for subgroups of examinees. An example of content bias against girls would be one in which students are asked to compare the weights of several objects, including a football. Since girls are less likely to have handled a football, they might find the item more difficult than boys, even though they have mastered the concept measured by the item (Scheuneman, 1982). An example of language bias against blacks is found in an item in which students were asked to identify an object which began with the same sound as "hand." While the correct answer was "heart," black students more often chose "car" because, in black slang, a car is referred to as a "hog." The black students had mastered the concept, but were getting the item wrong because of language differences (Scheuneman, 1982). Questions which might be asked to detect content, language, and item structure and format bias are given below.

Content Bias

There are at least five important questions that can be used to address content bias in a test:

- Does the item contain content that is different or unfamiliar to different DSI?
- Does the item measure what is taught in New York State high schools?

- Will members of DSI get the item correct or incorrect for the wrong reason?
- Does the content of the item reflect information and/or skills that may not be expected to be within the educational background of all test-taking examinees?
- Does the item content contain information that could benefit examinees of some DSI?

Language Bias

An item may be considered biased if it uses terms that are not commonly used state-wide, or uses terms that have different connotations in different parts of the state.

Three questions were generated for detecting language bias in a test:

- Does the item contain words which have different or unfamiliar meanings for DSI?
- Is the item free of difficult vocabulary?
- Is the item free of group specific language, vocabulary, or reference pronouns?

Item Structure and Format Bias

- Will any of the item distractors be unusually attractive to members of DSI for cultural reasons? (For example, some words may have different meanings in the first language of some of the examinees.)
- Are there any flaws in the items to which members of DSI are differentially sensitive?

- Does the item contain any errors or clues that make the various answer choices unequally attractive to members of DSI?
- Does the explanation concerning the nature of the task required to successfully complete the item tend to differentially confuse members of DSI?
- Will any of the distractors draw a disproportionate number of members of DSI? Are there flaws in the item that cause one or more options of the item to be attractive to members of DSI?
- Are clues included in the item that would facilitate the performance of one group over another?
- Should "I don't know" be included as an answer choice to prevent disproportionate amounts of guessing?
- Will the "correct or "best" answer change for different DSI?
- Will the use of a "negative" cause differences in performance?
- Are there any inadequacies or ambiguities in the test instructions, item stem, keyed response, or distractors?
- Does the format or structure of the item present greater problems for students from some background than for others?

Test Time Limits

Recently researcher have determined that restrictive time limits can be a source of bias on test items appearing late in a test. Judges, therefore, might be asked about the suitability of the time limits. Clearly, though, test speededness as a source of item or test bias is best assessed empirically.

Sample Item Review Form

What follows is an item bias review form. Alternately, an item fairness review form could be organized around the same issues and questions.

Item Bias Review Form[1]

This review form was designed to assist in the identification of items which may reflect bias against designated subgroups of interest (DSI). An item can be described as "biased" against a particular group of examinees when the characteristics or content of the item make the item more difficult for the group than would be predicted from a knowledge of the group's performance on other items in the test. This review form was also constructed to facilitate the identification of items which may show subgroups of examinees in stereotypical situations or eliciting stereotypical behaviors and emotions. Stereotyping of situations, behaviors, and emotions is undesirable although these misrepresentations often do not impact on group test performance.

At the top of the next page, please print your name, the data, and the test number. First, read through the review form to become familiar with the questions. Designated subgroups of examinees are referred to as DSI throughout. Next, read each test item in the test, and answer the questions concerning bias. Use "Y" for YES, "N" for NO, and "U" for UNSURE. On some occasions, a question may not be relevant for the test item. When this situation arises, indicate "NA" for NOT APPLICABLE. When you feel an item is "biased," beside the test item in the booklet please explain your response.

---

[1]Prepared by Ronald K. Hambleton and H. Jane Rogers from the University of Massachusetts at Amherst.

Item Bias Review Form

Reviewer's Name: _____ Date: _____ Test: _____

Stereotyping and Inadequate Representation	Item Number									
Does the test item										
1. contain material which is inflammatory, controversial or emotionally charged for members of DSI?										
2. contain language or material which is demeaning or offensive to members of DSI?										
3. portray members of DSI in situations that do not involve authority or leadership?										
4. depict members of DSI as experiencing stereotyped emotions?										
5. depict members of DSI as having stereotyped characteristics?										
6. depict members of DSI in stereotyped occupations?										
7. contain biased or offensive art work?										

Reviewer's Name: _____ Date: _____ Test: _____

Sex, Ethnic, Cultural, Religious, and Class Bias	Item Number									
Does the test item										
8. contain content that is different or unfamiliar to to some DSI?										
9. measure what is taught in New York state high schools?										
10. reflect information and/or skills that may not be expected to be within the educational background of all examinees?										
11. contain information that could benefit examinees of some DSI?										
12. Will members of DSI answer the item correctly or incorrectly for the wrong reason?										
13. Does the item contain words which have different or unfamiliar meanings for different DSI?										
14. Is the item free of group-specific language, vocabulary, or reference pronouns?										
15. Will any of the distractors be expecially attractive to members of DSI for cultural reasons?										

Reviewer's Name: _____ Date: _____ Test: _____

Sex, Ethnic, Cultural, Religious, and Class Bias	Item Number									
16. Does the explanation concerning the nature of the task required tend to differentially confuse members of DSI?										
17. Does the item contain clues which could facilitate the performance of members of some DSI?										
18. Will the "correct" or "best" answer change for different DSI?										
19. Does the format or structure of the item present greater problems for students from some backgrounds than for others?										

Reviewer's Name: _____ Date: _____ Test: _____

Please check the appropriate box for each of the questions below:

Overall Judgments	Yes	No	Unsure
1. Does the test cover topics of interest and relevance to DSI where possible?			
2. Does the test as a whole represent DSI positively in non-stereotyped ways and settings?			
3. Does the test as a whole represent DSI in proportion to their incidence in the population?			
4. Is the test content balanced in terms of being equally familiar to all DSI?			

References

Berk, R. A. (Ed.) (1982). Handbook of methods for detecting test bias. Baltimore, MD: The Johns Hopkins University Press.

Chipman, S. F. (1988, April). Word problems: Where test bias creeps in. Paper presented at the annual meeting of AERA, New Orleans.

Lawrence, I. M., Curley, W. E., & McHale, F. J. (1988, April). Differential item functioning of SAT-Verbal reading subscore items for male and female examinees. Paper presented at the annual meeting of AERA, New Orleans.

Mellenbergh, G. J. (1984, December). Finding the biasing trait(s). Paper presented at the Advanced Study Institute Human Assessment: Advances in Measuring Cognition and Motivation, Athens, Greece.

Mellenbergh, G. J. (1985, April). Item bias: Dutch research on its definition, detection, and explanation. Paper presented at the annual meeting of AERA, Chicago.

Scheuneman, J. D. (1981). A new look at bias in aptitude tests. In P. Merrifield (Ed.), New directions for testing and measurement: Measuring human abilities, No. 12. San Francisco: Jossey-Bass.

Scheuneman, J. D. (1982). A posteriori analyses of biased items. In R. A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore, MD: The Johns Hopkins University Press.

Scheuneman, J. D. (1984). A theoretical framework for the exploration of causes and effects of bias in testing. Educational Psychology, 19(4), 219-225.

Schmitt, A. P., Curley, W. E., Blaustein, C. A., & Dorans, N. J. (1988, April). Experimental evaluation of language and interest factors related to differential item functioning for Hispanic examinees on the SAT-Verbal. Paper presented at the annual meeting of AERA, New Orleans.

Tittle, C. K. (1982). Use of judgmental methods in item bias studies. In R. A. Berk (Ed.), Handbook of methods for detecting item bias. Baltimore, MD: The Johns Hopkins University Press.